

Persistent identifiers for Canada: The case for a national strategy

Background

In an increasingly global and digital research environment, we need to locate and link entities such as researchers, articles, datasets, and software across multiple complex information systems. At the moment, this is too often done manually or by inference. As a result, administrative burdens grow as research scales, basic data errors multiply, and the research information systems that underpin funding decisions and research careers are increasingly unwieldy and expensive to operate. Scholars are wasting valuable time constantly re-entering the same information; research analysis is costly, inefficient, and inaccurate.

Persistent identifiers (PIDs) are part of the solution. They can improve this situation by associating people (eg, researchers), places (eg, their organizations), and things (eg, research grants, outputs, and more) with a unique numeric or alphanumeric code (the identifier). PIDs are linked to structured information (metadata), which both describes the entity they identify, and also links to other PIDs and their associated metadata.¹ Ultimately, a PID is a long-lasting reference to a digital entity – a label which gives a unique code to a *person, place, or thing*.

Without the unique identification of researchers and outputs across the many systems used across the research ecosystem, it is impossible to accurately link grants, publications, data, and other activities to the right researchers and organizations – or to understand the impact of scholarship now and in the future. The Federal Government's 'Roadmap to Open Science'² highlights the need for agreement on the tools and infrastructure to underpin the transition to more open research, and emerging departmental Open Science Action Plans explicitly call for the adoption and implementation of PIDs.³

PIDs are a solution to many costly and complex information challenges facing the Canadian research ecosystem. They are a critical precondition for efficient, effective research in the modern world, underpinning sustainable and interoperable research data/information management systems.

¹ <https://www.jisc.ac.uk/guides/rdm-toolkit/identifiers>

² I. Government of Canada, 'Roadmap for Open Science', Feb. 2020. Accessed: Feb. 10, 2023. [Online]. Available: <https://science.gc.ca/site/science/en/office-chief-science-advisor/open-science/roadmap-open-science>

³ Open science action plan: response to the Government of Canada Roadmap for Open Science. Jun. 2021. Accessed: Mar 14, 2023. [Online]. Available: <https://nrc.canada.ca/en/corporate/planning-reporting/open-science-action-plan-response-government-canada-roadmap-open-science>

Digital Object Identifiers⁴ (DOIs) for publications and Open Researcher and Contributor Identifiers⁵ (ORCID) for those performing or supporting research, are already widely used. Like research itself, the organizations that register and support these PIDs are global and community-driven. On the other hand, research funding and policy are conducted largely at the national level, as governments and funders have a responsibility to develop and execute strategies that meet the needs of the communities they serve.

A national PID strategy will make the benefits that PIDs offer available to all Canadian scholars and institutions, as well as enabling Canada to advocate for its national needs and voices in the global conversation.

Global context

A number of countries are already developing national approaches and PID policies. For example, the August 2022 memo from the US Office of Science and Technology Policy (OSTP) on ensuring access to federally funded research⁶ recommends the use of ‘digital persistent identifiers’ for scientific and research and development awards, research outputs, and researchers themselves. The Research Data Alliance hosts a working group on national PID strategies,⁷ which has collected case studies from the UK, US, Australia, Netherlands, Finland, Peru, Brazil, Colombia, and South Korea, as well as Canada. The European Open Science Cloud has developed a specific policy on PIDs,⁸ and the 2021 G7 research compact⁹ commits member states to work together to improve “...*the availability, sustainability, usability and interoperability of research data, technologies, infrastructure and services*”. PIDs are also a vital component of the UNESCO recommendations on open science.¹⁰

The value of PIDs

Comprehensive adoption and use of PIDs can lead to significant savings in the time and money spent on common research administration tasks through automation of tasks and data re-use, for example, by

⁴ <https://www.doi.org/>

⁵ <https://orcid.org/>

⁶ The OSTP memo on equitable access can be found here: <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf> (accessed February 2023)

⁷ More information about the RDA working group on national PID strategies can be found here: <https://www.rda-alliance.org/plenaries/rda-18th-plenary-meeting-virtual/aligning-and-coordinating-national-pid-strategies> (accessed February 2023)

⁸ European Commission. Directorate General for Research and Innovation. and EOSC Executive Board., A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC). LU: Publications Office, 2020. Accessed: Feb. 10, 2023. [Online]. Available: <https://data.europa.eu/doi/10.2777/926037>

⁹ ‘G7 2021 Research Compact (PDF, 356KB 2 pages’, Group of 8, Russia. [Online]. Available: <https://policycommons.net/artifacts/1592313/g7-2021-research-compact-pdf-356kb-2-pages/>

¹⁰ <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>

auto-filling researcher CV information stored in their ORCID record.¹¹ These time savings translate to reduced administrative and institutional costs, but the value they bring goes beyond dollars and cents. PIDs have the power to support or enable critical research policy priorities.

Understanding the research landscape

PIDs expose connections and relationships between entities, such as linking individual grants to specific research outputs in an automated way without creating undue bureaucratic burden. At scale, this enables more strategic management of a national research portfolio, helping funders to understand the impact of past investments and how to best target future support.

Open research

Open PIDs are, as noted above, widely recognized as a central component of a successful transition to open research. As well as maximizing Canada's global scientific and economic impact and position, widespread PID adoption will also help ensure equitable access to information, as exemplified in the National Research Council's support for the use of PIDs¹².

Managing research data

Research data represents a tremendous source of future value. PIDs help make individual datasets visible, and also capture the information needed to enable the creation of new insights from old data, such as relationships to software, instruments, facilities, and researchers. This is why PIDs are critical to the FAIR (Findable, Accessible, Interoperable, and Reusable) principles¹³ for research data management.

Research integrity

PIDs enable the capture of information about the activities and context of knowledge creation early in the research lifecycle. They make it easier to understand current research activities and detect issues before they are published and distort the scholarly record. Post-publication, if issues are detected, PIDs make it easier to track flawed outcomes back to their source, and ensure that the underlying data or tools can be corrected or retracted.

¹¹ See, for example: Brown, Josh, Jones, Phill, Meadows, Alice, & Murphy, Fiona. (2022). Incentives to invest in identifiers: A cost-benefit analysis of persistent identifiers in Australian research systems. Zenodo. <https://doi.org/10.5281/zenodo.7100578> and Brown, Josh, Jones, Phill, Meadows, Alice, & Murphy, Fiona. (2022). Revised cost-benefit analysis for the UK PID Support Network (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.7356219>

¹² <https://nrc.canada.ca/en/corporate/planning-reporting/open-science-action-plan-response-government-canada-roadmap-open-science>

¹³ M. D. Wilkinson et al., 'The FAIR Guiding Principles for scientific data management and stewardship', *Sci Data*, vol. 3, no. 1, p. 160018, Dec. 2016, doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).

The case for central support in Canada

In 2021, acting for the Canadian Persistent Identifier Advisory Committee (CPIDAC), the Canadian Research Knowledge Network (CRKN) in partnership with the Digital Research Alliance of Canada, commissioned a research report¹⁴ into the current state of PID adoption, assessing opportunities for growth in, and barriers to, adoption in Canada.

The report found significant, cross-stakeholder support for the further development and adoption of PID infrastructures in Canada, with an existing strong community of PID experts. The Canadian research community overwhelmingly agrees that PIDs are necessary to satisfy a range of Canadian policy needs, and that a shared action plan is essential to maintain the technology, create and support communities of practice, offer education about best practices, and develop consensus around implementation plans and roadmaps that can be communicated to commercial development partners.

Increasing the adoption of PIDs will require collective action. PIDs represent an information network and, like all networks, the value increases as more people participate. For small, specialist, regional, or community-centred organizations, the initial cost of PID implementation may be a barrier. This risks the development of information inequality, with researchers and support staff at less well-funded institutions missing out on the benefits of PIDs, and consequently weakening the whole network. A centrally supported and resourced strategy is needed to address this risk and ensure equitable access to this critical toolkit for all.

Canada will benefit from a strong foundation for this work. The CPIDAC is well placed to take a leadership role in developing national consensus around suitable standards, workflows, best practices, and governance approaches. While PID providers already collaborate in a number of ways, they are now also recognizing the need to engage with national PID initiatives. A coherent and united national voice in this process will ensure that Canada has the opportunity to shape global priorities and to help PID and infrastructure providers better understand, and meet, the needs of the Canadian research system.

With the right support and investment, a Canadian national PID strategy will guarantee participation across the sector, enrich the level of interoperability and information exchange for all stakeholders, and generate new possibilities for Canadian research.

¹⁴ Brown, Josh, Jones, Phil, Meadows, Alice, and Murphy, Fiona, 'Towards a national PID strategy for Canada - Vers une stratégie nationale sur les PID pour le Canada', Zenodo, Oct. 2022. doi: [10.5281/ZENODO.7217469](https://doi.org/10.5281/ZENODO.7217469).